

ADITYA SHARMA

(778)-917-8028 | asa498@sfu.ca | [LinkedIn](#) | [Portfolio](#) | [Github](#)

CORE SKILLS:

Data Engineering & Processing: Dbx Lakehouse, Apache Spark, PySpark, Delta Lake, SparkSQL, SQL, Databricks SQL, MLflow
Business Intelligence & Analytics: Power BI (Fabric), DAX, Tableau, KPI Design, Dashboarding, Data Storytelling, Model Evaluation
Tools & Ecosystem: Apache Airflow, Scikit-learn, Git, GitHub, VSCode, Jupyter Notebooks, Docker (basic), Agile, Scrum, Data Modeling, Query Optimization, Partitioning, Indexing

CERTIFICATIONS:

Databricks Data Engineer Associate Certification (Issued: March, 2025)

PROFESSIONAL EXPERIENCE:

BIG DATA SUPPORT SPECIALIST

Sep 2023 – Apr 2024

Simon Fraser University | Vancouver

- Designed, implemented, and maintained scalable ETL pipelines using Python, Apache Spark, and Google BigQuery, processing millions of records with 99.9% uptime.
- Automated data ingestion from SFTP servers and REST APIs, enhancing real-time data processing capabilities by 30%.
- Developed and optimized SQL-based schemas for internal data warehouses, improving query efficiency by 40%.
- Monitored and troubleshooted operational issues in ETL workflows, ensuring data consistency and minimal downtime.

DATA ENGINEER INTERN

Sep 2020 – Jul 2021

Indian Council of Medical Research | New Delhi

- Built and optimized ETL pipelines with Python and SQL, processing medical data for large-scale epidemiological studies.
- Designed data models to improve vaccine efficacy analysis by 30%, ensuring compliance with data governance policies.
- Resolved data pipeline failures by analyzing root causes, reducing incident response time by 25%.
- Collaborated with stakeholders to create dashboards and reports, driving actionable insights for public health strategies.

JUNIOR DATA ANALYST

Nov 2019 – Dec 2019

NuWave Solutions pvt Ltd | New Delhi

- Optimized customer segmentation by analyzing 5 million transaction records using K-Means clustering in Python, leading to a 20% increase in targeted marketing effectiveness and customer retention.
- Enhanced demand forecasting with an ensemble regression model in Python, achieving 92% sales prediction accuracy while ensuring data confidentiality and integrity during processing.

PROJECTS:

ML-Powered Bank Campaign Optimization Using Databricks Lakehouse & Power BI

Jan 2025 – Mar 2025

- Built a complete ETL and ML pipeline in Databricks Lakehouse using the Medallion Architecture (Bronze/Silver/Gold) to identify high-conversion leads from 41K+ customer records in the UCI Bank Marketing dataset.
- Developed a Gradient Boosted Trees model using PySpark & MLlib achieving AUC = 0.864, helping simulate campaign outcomes based on lead probabilities.
- Designed a dynamic Power BI dashboard with interactive KPIs, funnel visualizations, and threshold-based customer targeting capabilities to support data-driven marketing decisions.
- Delivered a modular pipeline leveraging feature engineering, probabilistic modeling, and big data processing, enabling marketing teams to prioritize resources based on predictive analytics.

Data Driven Customer Segmentation and game recommendations for SFU

Jan 2023 – Apr 2023

- Developed an end-to-end ETL pipeline for web scraping, cleaning, and storing gaming data in BigQuery, which resulted in 89% recommendation accuracy and 40% improved student engagement.
- Performed feature engineering using SQL and dimensionality reduction with PCA, achieving 89% accuracy in game recommendations and enhancing predictive model performance.
- Developed a decision tree classifier using scikit-learn to forecast game success with an 84% F1-score, and created interactive visualizations using Tableau, which increased stakeholder engagement by 40% and revealed the impact of game genre and developer on success.

EDUCATION:

Master's of Computer Science in Big Data | Simon Fraser University, B.C

Sep 2022 – Apr 2024

Bachelor's in Technology in Computer Science | SRM Institute of Science and Technology

Jul 2018 – May 2022